CASTS Paper - Reviewer Comments

Reviewer 1 Comments

The paper outlines an easily accessible and quality-controlled product from multiple sources. It is explained well where data comes from and why there is a need for this data product in addition to the national database of the Canadian Integrated Ocean Observing System.
With the impact of decreasing ice sheets of Greenland getting more and more important, it is crucial to compile such datasets to build a strong community effort around it.
One of the purposes described are using it to review the changes in the ocean climate of Atlantic Canada. The paper conducts an initial analysis and finds that there are strong changes present in the Northwest Atlantic Ocean including decadal fluctuations. The authors make it clear that this is not an in-depth study of the area. The analysis done here is well suited for preliminary results. The analysis demonstrates that the dataset has potential.
This dataset also incorporates an older database, Climate, which was put offline in 2010 – through this dataset, they make the data accessible again. Not all data were quality controlled beforehand and the authors performed a separate quality controlled which is well described. This adds further value even if some of the data is available elsewhere. Other quality control algorithms are correctly cited and described where possible.
The manuscript is also transparent about known issues, like lack of metadata or no negative sign for temperature measurements. The format of yearly NetCDF files binned by pressure is well chosen as a easily digestable format.
I only have minor comments listed below.

Response

We appreciate the helpful comments from the reviewers and editors. We believe that the manuscript is now much stronger after addressing their concerns. The reviewers comments are listed below with the italicized responses following. Changes to the manuscript and their associated line number are highlighted in blue text.

1. L39 Are there any considerations when comparing century old data to more recent one? I am aware that for each individual datasets you discuss pre-screening and data issues, but perhaps it would be interesting to mention changing data quality here.
    a. **Response**
    We do not currently perform any QA/QC steps relating to potential bias between instrument types and the time they were taken beyond those which were completed internally by the relative sources. We hope to address this shortcoming soon and have highlighted this future step in L 653. We also wanted to allow the end-user flexibility, so that they can choose whether or not to apply their own corrections to a specific instrument (e.g. XBTs). A specific instrument can be easily isolated using the instrument_ID_manual variable. See also comments 2 from Reviewer 2 below.

2. L120 What an annoying issue to encounter! Was this an issue with these specific instruments or measuring protocols? Perhaps a little information would reassure

someone working with the data that all of these issues have been caught and it's not a problem for further datasets.

   a. **Response**

      We agree with reviewer 1 that further reassurance should be given to the reader by clearly stating all problem profiles were removed following a thorough visual inspection. L 132 has been edited to address this issue. "These casts were manually removed from the data product after an exhaustive visual inspection of the meta data (this problem only affected data that originated from one source)."

3. L145 The way the automatic flagging sounds good - but I'm just wondering if there were any additional manual checks after flagging? Perhaps there were too many flagged profiles, but in that case it would also be interesting to hear how much was flagged.

   a. **Response**

      No manual check was performed during this QA/QC step. We also agree with reviewer 1 that the number of flagged/removed profiles for each QA/QC step should be provided, and not just for L 141. Additional information has been added on L 141 and L 156 regarding the NAFC - Oceanography methodology and L 351 and L374 for the combined data product. Example text for L 141 is the following: "(3032 profiles removed)".

4. L299 Are the headers of the resulting product the same as in this step?

   a. **Response**

      Yes, the headers of the resulting product are consistent.

5. L303 Great to provide this example command and also the example header above. Perhaps most users don't need aid with opening data, but I think this is useful for demonstrating how easy to access the dataset is.

   a. **Response**

      Thank you.

6. L314 You said you expect that the Climate quality control is better – that is probably right, but it would be great to hear more about why you think it's better or how much worse the other data is

   a. **Response**

      We agree with reviewer 1 that further clarification should be provided regarding the decision to prioritize profiles from Clime over NAFC-Oceanography sourced profiles. Further clarification is added in L 337. "Given that casts sourced from Climate went through an exhaustive quality control process (Gregory 2004), while other historical sources such as NAFC-Oceanography did not, we therefore gave priority to the Climate data set."

7. L329 Why exceeding the depth by 200 m - and not strictly that depth? Is this to account for some uncertainties in the GEBCO dataset?

   a. **Response**

UNESCO 2010 2.11 Bottom test looks for profile depths that are 50m greater than the reported GEBCO bathymetry in the area. This makes our choice of 200m conservative by comparison. This depth was chosen to allow for the possibility of larger depth errors for locations along the shelf break where the slope is greater and in remote northern areas of the domain where bathymetry uncertainty is higher. It was found that the 50m depth cutoff used by UNESCO 2010 was too strict for these locations. An additional sentence has been added to explain this choice L 354. "The Intergovernmental Oceanographic Commission (2010) recommends a 50m cut-off to allow for the possibility of slight variations in profile location along the shelf break without entirely removing the profile. After inspection of the data, we instead chose a 200m depth cut-off to account for larger uncertainties in bathymetry in some areas of the domain (slopes, seamounts and canyons, especially in the northern part of the domain)."

8. L343 As before, it would be good to hear if manual checks were at some point performed
    a. **Response**
    We agree with the reviewer that the paragraph should specify whether a manual check was performed at this point, similar to the NAFC-Oceanography specific outlier check. An additional sentence was added at L 371. "If a temperature or salinity measurement was outside the mean plus or minus 40 times the bootstrap standard error, then the measurement was identified as an outlier. No manual comparison check was performed."

9. L360 This overview is useful, but perhaps it can be moved to the beginning of the section, after which the steps are described?
    a. **Response**
    While we agree that the methods summary bullets being located at the beginning of the methods section would be useful, we feel that it makes more sense to include it at the end. Currently, the numbering of the list in Section 4.8 does not reflect the Section 4 numbering. We therefore decided to keep it at the end of the section so as to not confuse the reader.

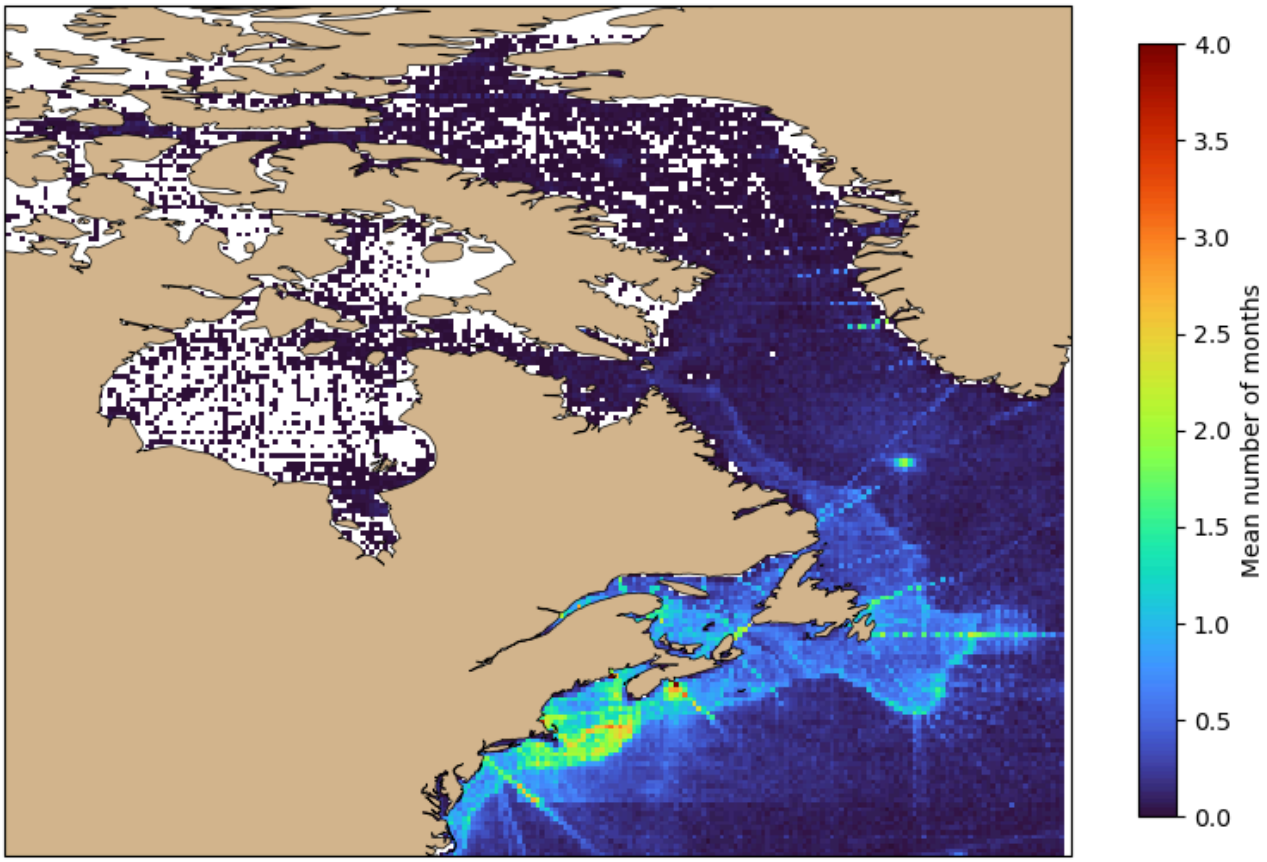
10. L414 It's ARGO floats, not ARGOS
    a. **Response**
    The typo has been corrected to Argo.

11. L414 Can you clarify if ARGO floats are included in the dataset? I know that you mentioned "profiling floats" before in section 3, but perhaps it would be good to address the ARGO program separately
    a. **Response**
    We agree that further clarification should be added on whether Argo profiles are included in CASTS. Argo is now considered a standalone source and a separate section has been added to the methodology section starting on L 299. Figures have also been updated accordingly.

12. L575 I mentioned this before, but for data stretching back to 1920 I would like to hear a bit more about differences in data quality or biases
    a. **Response**
    We agree with reviewer 1 that additional information should be added concerning the QA/QC of early data from sources such as WOD. An additional sentence has been added to the WOD section to address this on L 94. "WOD temperature and salinity profiles undergo a suite of quality assurances and controls, including data formatting, measurement position and time, gradients, and variable ranges (Mishonov, 2024). This gives us confidence to use and compare measurements taken over multiple time periods."

13. L600 You have several recommendations before you say you're updating the dataset in the coming years. Are these recommendations simply recommendations for further datasets or are they future plans for CASTS?
    a. **Response**
    We see these recommendations as our list of future plans for CASTS to improve upon the data product with each new iteration. A clarifying sentence has been added on L 657. "We hope to add these recommendations in future iterations of the CASTS data product."

14. Figure 5 You previously mentioned in the text e.g. how many months on average are available in a given dataset that you included. Perhaps it would be interesting to add information about seasonal bias. This could be a map similar to this one, which shows total number of profiles, but instead e.g. average number of months covered in a year or a similar indicator. Maybe this would look too similar to the total number of profiles, however.
    a. **Response**
    A sample figure was created to see if an additional figure showing the average number of months available per year would be beneficial to the paper (see below). However, we have found that it is very similar to the figure showing the total number of profiles and does not add much novel information to the paper. We have therefore decided not to include the figure.



15. Figure 8 I think for the purpose of this plot it is still okay, but note that these rainbow-style colourmaps are not perceptually uniform (see e.g. Thyng et al. 2016 at http://dx.doi.org/10.5670/oceanog.2016.66)
    a. **Response**
    While we are aware that the rainbow-style colourmap is not perceptually uniform, the colour map was chosen to highlight the CIL layer clearly. The figure caption has been updated to inform the reader of this choice. "The colour-map used in the first row is not perceptually uniform and was primarily chosen to highlight the CIL layer clearly (i.e., light purple colour)."

Reviewer 2 Comments

This paper describes a chronologically assembled quality-controled data set of hydrographic (T-S) profile data mostly located along the shelves and continental slopes off maritime Canada (but extending wider).

General Comments:

1. The title also mentions the Eastern Arctic. This is a very Canadian-oriented designation for that part of the Canadian Arctic (including Baffin Bay), and I am not sure how much it is used elsewhere. Here, the description/investigation stops at Davies Strait: what is referred to as Canadian Atlantic shelf… (even though the data coverage is 35-80°N and 42-100°W; thus also includes for example Flemish Cap, but not the slope region east of it). Considering the focus and the examples provided to illustrate the products, I am not so sure that 'Eastern Arctic' should be part of the title.
    a. **Response**

    We agree with reviewer 2 that the usage of the term "Eastern Arctic" in the title may be confusing. We have therefore decided to change the title of the paper to "More than a century of oceanic hydrographic observations reveal profound climate-related changes in the Northwest Atlantic Ocean and its Arctic Gateways".

2. A large compilation of data is assembled from different sources, with a rather logical selection and validation process. There are however no correction of potential biases (for example, for MBTs or XBTs), and some data types are skipped (mooring, gliders, and possibly seal-borne probes, but that is not clear for the latest with subsets are included). THe effort is very valuable and worth been published in ESSD. However, I still have some concerns:
    a. **Response**

    There are two valid points in this comment and we will address them separately.

    First, it is indeed true that older instruments such as XBTs and MTBs are known to have some instrument-specific bias and that corrections have been recommended by previous studies. One difficulty to apply those corrections is that they depend on the manufacturer of the sensors (Cheng et al., 2016), and that the meta data containing this information have been lost for older data. Our approach here is thus to leave these corrections to the discretion of the user rather than trying to impose a specific correction. Because XBT and MBT data are identified with the variable instrument_ID_manual, the user can easily remove or correct those data using custom correction algorithms. We have added a paragraph at the end of Section 5 to better acknowledge this caveat: "We also note that older instruments, such as XBTs, MBTs or BTs are known to have instrument-specific biases (e.g., for temperature and depth) and that corrections have been recommended by previous studies (e.g., Boyer et al., 2016, Cheng et al, 2016, 2018). Those biases, for example positive for temperature on the order of 0.13° - 0.15°C for analog systems and 0.01° to

0.07°C for digital systems (e.g. Emery et al. 1986; Cowley et al. 2013, Cheng et al., 2016) are extremely important for the calculation of global heat budget of the ocean (Ishii et al., 2009). However, they may be negligible for some of the coastal observations presented in this study where interannual fluctuations are large (several degrees). As the recommendations for those bias corrections are rapidly evolving, and because some of the meta data for those corrections (e.g., instrument manufacturer) is not available, our strategy for this data product was to leave those observations uncorrected. The user can apply a correction at its own discretion or simply discard those observations since they are identified with the variable "instrument_ID_manual".

Second, moorings, gliders, and seal-borne probes were specifically removed from sources and not included into CASTS due to difficulties with integrating the data sources with traditional temperature-salinity profiles. An additional sentence explaining this choice has been added to the paper on L 55. "CASTS focuses specifically on temperature-salinity profiles, while removing glider, mooring, and seal-borne probes due to data formatting difficulties (e.g. high temporal frequency measurements, non-vertical profiles, etc.) when integrating these specific data sources. TS data from these sources may be added at a future date."

3. - the presentation for the data is as time series mostly from a few best sampled stations. This is restrictive, with thus a focus onthe temporal sampling, but not the spatial sampling. Although the finally-poduced data set is indeed chronologically sorted, what this brings for other parts of the whole domain is not really discussed. However, ranges of variability (or standard deviations) are used in a spatial way to select/vlidate the data, which are not presented.They should at least be included in supplementary material (I could not open the Supplementary mterial, which seems only to accept Winzip as a tool, and not other tools)
    a. **Response**
    While we mostly present temporal data here, the CASTS data product can also be used to produce spatial output (e.g. CABOTS, https://www.frdr-dfdr.ca/repo/dataset/c713188f-3e24-4399-9dde-94331ab0a704). Within the extent of this paper, we have decided to focus on a few highly sampled areas to illustrate the temporal extent of the CASTS data product. We will add the climatology files to the supplementary material as suggested along with an example outlier profile figure.

4. - Furthermore, the different times series presented (with a clear focus on the annual (or seasonal) means (due to the sampling and natural higher frequency variability, even after removing an average seasonal cycle)) are not associated with uncertainty estimates. There is one attempt to create an index based on spatially-distributed data (the CIL on the GB), but it is not clear from the presentation how it is estimated (how the spatial information is taken into account, for instance and at which level of the averaging...). The huge area over which it is estimated (extending right over the Flemish

cap to the east and to Newfoundland and station 27 to the west) is not clear. I wondered reading it whether station 27 would not include a strong percentage of the winter data. Is there an assumption that it has no spatial structure (at least statistically). But how true is it over such a large area (and what about CIL area indices?)

   a. **Response**
   We agree with Reviewer 2 that a potential spatial bias should be addressed in the CIL figure. We have therefore decided to restrict the CIL spatial area to around Station 27 to ensure that spatial bias is minimized. A new updated figure and updated section is now available.

5. Finally, the choice of the time series is at times a little bit peculiar, and it is not completely clear what the criteria were for selecting these illustrating examples (in section 6) (see specific comments in the detailed comments).
   a. **Response**
   Addressed in the following comments.

Detailed comments:

1. Introduction: l. 21, connection of sea ice loss in the Arctic and Greenland Ice sheet loss to the freshening in the NASPG rather controversial… (accumulation for example recently of liquid freshwater in the Arctic after sea ice melt suggesting that net budget of the export of fresh water (either liquid or in sea ice) is not that straightforward…)
   a. **Response**
   We agree with reviewer 2 that the uncertainty surrounding the freshening in the NASPG research should be highlighted. A sentence has been added to reflect this on L 24. "The extent to which the recent freshening can be explained by sea ice melt and glacier runoff is still an area of active research (e.g., Yashayaev 2024).

2. The later mention on that page of the confluence of water masses for the large decadal or longer time scale variability in this region right, but on coastal areas (and for T) also with large continental influences.
   a. **Response**
   We agree with reviewer 2 that a distinction should be made between near-shore and offshore areas of the shelf when describing the physical forcings. A sentence has been added to address this point on L 40. "Areas of the domain close to shore are also being increasingly influenced by continental forcings such as river runoff (Haine et al 2015)."

3. l. 85: the collection from WOD is only for station data and MBT. Interestingly, not XBTs or pinnipeds. Is it that they are kown to be better validated elsewhere? What about gliders (surveys known off Halifax or south-east Newfoundland) (I see it mentioned in 3.3, with the mention that they were removed; data not calibrated?)?
   a. **Response**

We agree that further information should be provided regarding the instrument types included in the WOD source should be expanded upon. We have also increased the use of WOD as a source to include 2008-2018 data so as to decrease our reliance on BIO-OMM. Separate information is provided for what instrument sources are used during these years.

The glider comment is answered in Reviewer 2, general comments number 2.

4. l. 88: add year of the occupation of the Challenger Halifax sections (I guess 1873). L. 89 'missing years between 1873 and 1910' is it 'all years' or 'some of the years'
    a. **Response**
    We agree with reviewer 2 that the exact years which are missing should be stated. The sentence has been added in L 88. "The following years have no data available; 1874-1876, 1878-1882, 1888, 1890-1894, 1896-1897, 1899, 1901, and 1904."

5. l. 95: for the NAFC database, among other countries mentioned, would the UK have been doing surveys in the area of Newfoundland-Labrador been a dominion (1909-1949)
    a. **Response**
    We agree with reviewer 2 that other countries which have significantly contributed to the database should be highlighted. The sentence L 106 has been edited. "The NAFC regional data center has historical observations collected by Canada, the United Kingdom, France, and other countries."

6. l.. 279: strange to use keyword CTD for data in 1912-2023 (and title of 3.8 is supposed to be 2002-2023, as also explained l. 282)
    a. **Response**
    We agree with Reviewer 2 that the use of the keyword CTD while specifying a time range starting in 1912 is strange. We have changed the title and accompanying source to the recommended 2002-2024 as the reviewer suggests on L 288.

7. In tests, wondering how many profiles removed by the 4.2 test. The test on T-S profiles with 0 standard deviation if more than 3 measurements is a bit extreme. Depending on reported resolution, season, range of depth, this could be physical
    a. **Response**
    We agree that further information should be provided on how many profiles are removed during most of the steps. Additional sentences have been added to the methodology stating the number of profiles removed during each step. The number of profiles removed during the initial combining of all sources is omitted due to the high amount of overlap between sources.

8. 4.4 the depth issue. There was a published paper using seal data off south-east Greenland showing that the maximum depths reached by the seals were some times

larger than the bottom depth reported in GEBCO, indicating small scale errors in GEBCO (but I don't recall whether this was exceeding 200m, and possibly bathymetry much better on Labrador – Newfoundland shelves (and certainly better in the southern part of the domain!). Are these errors of depth of cast mostly associated with XBTs, or also station/CTD casts (in which case a position error is likely, such as sometimes happens if data reported initially in degrees, minutes and not decimal degrees. With XBTs, it is unfortunately common that cut of profile when reaching the bottom is not correctly applied (I am sorry I am not able to find the reference)
Here too, wondering what is the number of profiles removed by the check and which type.

   a. **Response**
      We address this in comment 7 from Reviewer 1. We also now provide the number of profiles removed during this step.

9. l. 414: 'such as Argo' (not 'ARGOS')
   a. **Response**
      Corrected.

10. For the illustration of time series, l. 436: why use the CIL Minimum temperature at Rimouski station, instead of CIL thickness layer (as is used for the Grand Banks surveys?). I am also wondering about the choice of showing the vertically averaged T and S. Maybe this should be argued a little. Is there a little bit more time coverage if the focus was on specific layers and not the vertically averaged (for example, near surface layer maybe a little more sampled at some of the sites?). Is the variability in the different layers well correlated? (well, this is discussed a little through the upper layer stratification (0-50m), but there does it make much senses to present its annual average, as it presents a very large seasonal cycle; should the focus be more seasonal for that parameter?)
    a. **Response**
       The reason for the CIL minimum temperature or thickness not being included is because the total depth at Shediac Valley station (84m) is usually shallower than the deep limit of the CIL at nearby locations (i.e., the CIL extends all the way to the seafloor). This makes the calculation of the CIL minimum temperature or thickness impossible. A sentence has been added at L 477 to justify this omission. "The CIL minimum temperature was not calculated at the Shediac Valley Station due to the station depth being shallower than the CIL maximum depth at nearby locations."

11. Also, in the explanation of Fig. 8 on lines 441-445, I am not exactly sure of what is done, as the figure 8 suggests that this is the spatial average, corresponding to the box on Fig. 1. However, there is some spatial variability through the region, and how is it is taken into account. Is the time series analysis done first at each grid point and then averaged spatially, or is the assumption that there is no spatial structure of the CIL depth through this large region. If so, what justifies it?

a. **Response**
Please see Reviewer 2 general comment 4.

12. A bit difficult in section 6 to follow the figures. They are all cited at the beginning of section-, but after discussed sequentially.
    a. **Response**
    We agree with reviewer 2 that references to the time series figures in section 6 should be improved. More references to the two figures have been added to the section.

13. l. 450: station 27 representative of the NW Atlantic as a whole. To me rather far-fetched. Maybe OK for the shelf conditions, but not so close with NW Atlantic off-shelf, even rather close by…
    a. **Response**
    *We slightly modified the statement to emphasise that Station 27 is not representative of the NW ?Atlantic as a whole, but rather of its interannual *climate* fluctuations. This statement is backed up by several studies that found that Station 27 observations track the large-scale climate and its interannual variations very well (L 494):*
    *Its local ocean parameters are *considered a good proxy* for the interannual fluctuations of the NL Shelf climate and the *climate of the* NW Atlantic as a whole (Petrie et al., 1991, 1992; Colbourne et al., 1994; Drinkwater, 1996; Han et al., 2015; Cyr and Galbraith, 2021).*

14. l. 468: 'instead of melting ice from the Arctic…', could be more 'intensified export of freshwater'(freshwater can be transported either in liquid water or sea ice; with possible contribution of increased melt of continental ice', as the station is very close to shore, and could receive melt from Baffin Island or further north, as well as Hudson Bay; I am less sure for Greenland)
    a. **Response**
    We agree that more information should be provided about the potential source of the intensified export of freshwater. The following sentence has been added on L 511. "This pattern is consistent with other observations from the North Atlantic and may be a consequence of increased freshwater export from the Arctic since Station~27 is located downstream of the Davis and Hudson Straits."

15. l. 470: the pre-requisite of that statement is that there is near-surface intensification of these T and S signals. At that point, we do not know whether this is the case. I suggest to use 'could' instead of 'should'. Indeed, the results presented are not that coherent (see the weak stratification in the late 2010s).
    a. **Response**
    We agree with reviewer 2 and the sentence has been changed on L 514.

16. l. 505: I was also wondering whether there was a tendency to have a correlation between the Prince 5 and Halifax2 salinity anomaly time series. At least, it is the impression given by the figures (for the most recent period)
    a. **Response**
    We agree with Reviewer 2 that it is important to highlight the visual correlation between Prince-5 and Halifax 2 salinities in recent years. A sentence has been added on L 550. "Prince-5 and Halifax 2 salinity in the last 20 years have shown a strong correlation (0.50)."

17. l. 521: 'correlated with seasonally averaged St. Lawrence runoff'… (this is not shown, but I wondered whether the results cited from other papers are with the time series plotted: is it correct?)
    a. **Response**
    The paper (Galbraith et a., 2025 AZMP zonal report) includes the St. Lawrence runoff scorecard which is being plotted here. The paper also states that "stratification at Rimouski station and St. Lawrence runoff are correlated (r-squared value of 0.65 from 1991-2024)" which is included in the CASTS paper on L 565.

18. l. 533: how is it distinctly different from other local water masses, such as the continental slope water (I assume east of the Banks?). To which extent is it shelf winter water, maybe formed a little upstream and advected to the area sampled on the Grand Banks and its vicinity (the box plotted on Fig. 1 is rather large) (to some extent, this is commented after on l. 536, and even after on l. 540-541… and I suggest that all these comments and explanations on the CIL be grouped together.
    a. **Response**
    Please see Reviewer 2 general comment 4.

19. l. 538: I assume that it should be '… a strong seasonal signal, with temperature in the upper 50 m climatologically larger than 7°C…'
    a. **Response**
    We agree with reviewer 2 and have changed the sentence on L 570.

20. l. 558: the end of the sentence 'especially during… at shallower depths'. What does it refer to in the sentence? Maybe a separate sentence, indicating that the shallower depths have been more systematically sampled in recent years than earlier.
    a. **Response**
    We agree with reviewer 2 and have added an additional sentence explaining that more profiles are taken on the shelf (shallower depths) when compared to the open ocean. A sentence has been added on L 607. "A higher number of profiles are available at shallower depths on the shelf as that area has been more systematically sampled in recent years."

21. l. 566: I would skip ‘, whether through ice melt or river runoff’, which I find restrictive for the conclusions of Florindo-Lopez et al (2020) paper. Off course, this paper deals with interesting surveys of the Seal Island section, further north, and in particular the Coastal current near southern Labrador. It is not clear to me that this is the water mass most sampled at station 27, which I suspect also has a large inflow of the Baffin Bay branch.
    a. **Response**
    We agree with reviewer 2 and have removed the second half of the sentence.

22. On Fig. 1 box ‘Grand Banks’, goes quite far east on Flemish Pass, but does not cover the Grand Banks further south… I guess the choice is due to data cover.
    a. **Response**
    Please see Reviewer 2 general comment 4.

References

Boyer, T., Domingues, C. M., Good, S. A., Johnson, G. C., Lyman, J. M., Ishii, M., ... & Bindoff, N. L. (2016). Sensitivity of global upper-ocean heat content estimates to mapping methods, XBT bias corrections, and baseline climatologies. *Journal of Climate*, *29*(13), 4817-4842.

Cheng, L., Abraham, J., Goni, G., Boyer, T., Wijffels, S., Cowley, R., ... & Zhu, J. (2016). XBT science: Assessment of instrumental biases and errors. *Bulletin of the American Meteorological Society*, *97*(6), 924-933.

Cheng, L., Luo, H., Boyer, T., Cowley, R., Abraham, J., Gouretski, V., ... & Zhu, J. (2018). How well can we correct systematic errors in historical XBT data?. *Journal of Atmospheric and Oceanic Technology*, *35*(5), 1103-1125.

Ishii, M., & Kimoto, M. (2009). Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *Journal of Oceanography*, *65*(3), 287-299.